R E P O R T   R E S U M E S

ED 014 877   24   EM 004 010

PROGRAMMED INSTRUCTION IN THE INTACT CLASSROOM.

BY- GLASER, ROBERT AND OTHERS
PITTSBURGH UNIV., PA., LEARNING RES. AND DEV. CTR.

REPORT NUMBER CRP-1343-2   PUB DATE DEC 63
EDRS PRICE MF-$0.25 HC-$2.04   49P.

DESCRIPTORS- *GROUPING (INSTRUCTIONAL PURPOSES), *PROGRAMED INSTRUCTION, *PACING, *LEARNING MOTIVATION, *ACADEMIC ABILITY, TEACHING MACHINES, PROGRAM EVALUATION, CLASSROOM ENVIRONMENT, BALDWIN WHITEHALL SCHOOLS, PITTSBURGH, PENNSYLVANIA

VARIABLES INFLUENCING EFFECTIVE PROGRAMMED INSTRUCTION WERE STUDIED WITHIN THE INTACT CLASSROOM STRUCTURE. LATER STUDIES FOCUSING ON EFFECTIVE PROGRAMMED INSTRUCTION WILL APPROACH IT AS A MEANS OF INDIVIDUALIZATION OF INSTRUCTION. THIS SERIES OF EXPERIMENTS USES LINEAR PROGRAMS IN ARITHMETIC, SPELLING, AND GENERAL SCIENCE. FOR GRADE 1, ASPECTS STUDIED INCLUDE TEACHING MACHINES, TEACHER-PROGRAM ARRANGEMENTS, DAILY WORK-PROGRAM DISTRIBUTION, PREFAMILIARIZATION AND POST-LEARNING PRACTICE. FOR GRADE 4, EFFECTS OF PROGRAMMED INSTRUCTION ON REVIEW, ACCELERATION, AND CLASSROOM ENVIRONMENT ARE STUDIED. FOR GRADE 7, ENRICHMENT ACTIVITY, OVERVIEW OF MATERIAL, AND PREFAMILIARIZATION ARE KEY VARIABLES. FOR GRADE 9, EFFECTS OF A PROGRAM ON HIGH AND AVERAGE IQ GROUPS ARE STUDIED. ALTHOUGH THE AUTHORS WERE AWARE OF THE NECESSITY TO CONTROL FOR THE QUALITY AND SUBJECT MATTER OF PROGRAMMED MATERIALS, TEACHER CHARACTERISTICS, CEILING EFFECTS OF ACHIEVEMENT TESTS, AND CLASS DIFFERENCES IN ABILITY, THESE FACTORS WERE NOT WHOLLY CONTROLLED IN STUDY DESIGN, BUT THEY WERE CONSIDERED IN REPORTING RESULTS. GENERAL CONCLUSIONS ARE--EXTENSIVE VARIATION IN LEARNING RATE PREVAILS UNDER SELF-PACING CONDITIONS, PRETEST SCORES SHOW THAT MANY STUDENTS KNOW THE SUBJECT AND SOME ARE NOT READY TO LEARN, INTELLIGENCE IS RELATED TO PACE, STUDENTS REQUIRED TO LEARN MORE DO LEARN MORE, DIFFERENT TEACHER-PROGRAM COMBINATIONS IN SEVERAL GRADES DO NOT AFFECT ACHIEVEMENT. (LH)

PROGRAMMED INSTRUCTION IN THE INTACT CLASSROOM

Robert Glaser
James H. Reynolds
Margaret G. Fullick

Learning Research and Development Center
University of Pittsburgh

Project No. 1343
Cooperative Research Program
U. S. Office of Education

December, 1963

Programmed Instruction in the Intact Classroom[1]

Robert Glaser
James H. Reynolds[2]
Margaret G. Fullick

University of Pittsburgh

When the use of programmed instructional materials is studied in a school system, two different approaches to implementation can be considered. One approach considers the concept of programmed instruction as a means for individualizing the instructional process. The other approach conceives of the program as a reproducible teaching tool that can be used in various ways to improve the instruction of the class as a group.

The first notion, that of individualization, is a primary assumption behind the development of programmed instructional procedures. Ideally, programmed instruction is a means whereby the student can be provided with instruction on the basis of his particular requirements. A tutorial process is the analogy of the individualization process. The efficient tutor determines in detail the knowledge and skill that the student has prior to instruction; he then begins instruction assuming the competences that the student has shown. The instructional procedure is adjusted for the student by the tutor according

---

[1]The research reported herein was supported through the Cooperative Research Program of the Office of Education, U. S. Department of Health, Education and Welfare.

Appreciation must be expressed to the administrators of the Baldwin-Whitehall Schools, Dr. W. Robert Paynter, Superintendent; Dr. Warren D. Shepler, Assistant Superintendent in Charge of Instruction; Mr. J. Ernest Harrison, Director of Curriculum and Research; and the participating principals and teachers for their willingness and cooperation in formulating and carrying out the studies described.

[2]Now at the Institute for Human Learning, University of California, Berkeley, California.

to the rate at which the student learns, the kind of forward steps the student can take, and the kind of experiences which the student finds rewarding and motivating for effective attainment of subject matter mastery.

Current use of programmed instruction has far from attained the ideal of the individualization of instruction. At the present time, however, programmed materials and the concepts underlying them represent a step toward the provision of an individualized instructional environment for each student.[3] This is so to the extent that present programmed materials can permit the student to learn at his optimal rate and have the freedom to move ahead or catch up depending upon his mastery of the subject matter. When programs are used with such individualization in mind, they obviously necessitate restructuring of the intact classroom unit because different students in the class will require different instructional conditions and subject matter at different levels. Such reorganization is considered desirable by many school administrators but is a major problem for a school system where the unit of organization is intact class groupings and yearly grade-by-grade advancement.

Within the intact classroom structure, experimentation with programmed materials can be carried out by manipulating certain aspects of the classroom instructional procedure. In this way, the achievement of the class and changes in classroom teaching procedures become, respectively, the dependent and independent variables for study. For the most part, it is the use of programs in the intact classroom which is reported in this chapter. While this is not in keeping with the individualization assumption of programmed instructional concepts, it is likely that many school systems will first use

[3]This does not imply, in any sense, depriving the student of opportunities for socialization and interaction with his peers.

programmed instructional materials in intact classroom groups prior to the more extensive reorganization required for individualized instruction.

During the academic year 1962-63, programs which were available at certain grade levels were used in the Baldwin-Whitehall Schools in suburban Pittsburgh. The studies reported in this chapter represent attempts to study the use of programmed instructional materials within existing classroom structures in this school system. In the academic year 1963-64, studies will be performed which are oriented toward investigating the requirements for effective individualization of the instructional process.

The Questions Asked

The questions which were asked about the use of programs arose from primarily two sources: (1) variables studied in the psychologist's learning laboratory that suggested procedures for improving instructional effectiveness, e.g., the distribution of practice, and (2) problems arising from general teaching practices and educational requirements, e.g., the necessity for providing extended opportunities for learning. Sometimes both of these sources provide the background for a particular experiment.

Studies were designed to investigate the following kinds of questions:

Grade 1 - Can simple teaching machines be used in the classroom with young children beginning the first grade? What is the relative effectiveness of different teacher-program arrangements upon learning? What is the relative effectiveness of varying the distribution of daily work with the program? What is the effect of prefamiliarization and post-learning practice in the achievement resulting from programmed instruction?

Grade 4 - What is the relationship between intelligence and use of programmed instruction under certain conditions? How effective is programmed instruction for the review and acceleration of learning? What are the effects of classroom surroundings upon learning from a program?

Grade 7 - What is the effect of various combinations of programmed instruction and enrichment activity? Does prefamiliarization and an overview of material to be learned improve the effectiveness of programmed instruction?

Grade 9 - What is the effect of a program on high and average I.Q. groups?

Some of the questions listed above are only peripherally touched upon in this chapter and are reported in more detail in a larger technical report (Reynolds and Glaser, 1963).

Control Aspects

When specific studies are set up in an on-going school situation to answer these questions, a variety of variables must be considered which can influence the data obtained. The influence of these variables must be considered in interpreting the results of the studies or must be controlled in some way. The following aspects were of concern in the studies reported in this chapter.

(1) The quality of the programmed instructional materials. With the exception of one program constructed at the University of Pittsburgh, the programs employed were commercially available from reputable program publishers. These publishers provided some evidence that the programs were constructed according to good program development practices and were effective instructional instruments. This evidence was of an informal nature, since most program publishers at the present time do not provide manuals giving detailed program use

and validity data. To a large extent, this is a function of the newness of programs, and manuals similar to those accompanying nationally-standardized tests, containing the validation data obtained during the course of program development, are being made available. Furthermore, standard criteria which publishers can follow in the development of a program manual are being developed by national committees (Joint Committee on Programmed Instruction and Teaching Machines, 1963).

The degree of effectiveness of the various programs was not specifically known prior to use, and the efficiency and effectiveness with which they taught varied. The extent to which the effectiveness of a program interacted with the particular study being carried out is difficult to assess, and the differences in this variable of program quality was controlled only to the extent that some impression was available about their initial construction and subsequent development and use.

The common type of program on the school market is the linear program in which all students go through the same materials and no provision is made for "branching" sequences in which students are guided to different levels of material on the basis of their performance in the course of the program. All programs used in these studies were linear in format.

(2) <u>The subject matter</u>. The kind of subject matter taught by programs in the present studies was selected on the basis of (a) availability for the particular grade levels involved and (b) subject matter requirements in terms of student need and student-teacher-community acceptance as determined by the Baldwin-Whitehall school administrators. All of the programs were considered by teachers and administrators as representative of the subject matter normally taught at the grade level in which they were introduced. Of the eight commercial

programs used, six pertained to arithmetic or mathematics; the two other subject matters used were spelling and general science. This reflects the fact that a preponderance of programs available at the time were on mathematics topics and that this is a topic readily introduced into school systems in program form. The extent to which the greater number of programs in mathematics influenced the results of the studies carried out is again difficult to assess.

(3) Teacher characteristics. In all of the studies, the teacher participated to a greater or lesser degree in instruction in the subject matter involved in the program. As a result, differential teacher characteristics could influence the data obtained. However, since intact classes were used, teacher characteristics were controlled to the extent that whenever possible at least two different teachers were involved in each of the experimental conditions compared. This limited control had the effect of preventing any one experimental condition from dependence upon a single teacher. In addition, all teachers involved in the various studies were chosen on the basis of a positive (or at least non-negative) interest in trying out programmed instructional materials. Previous to classroom instruction, teachers participated in the development of the particular procedures to be used; one teacher for each study at each grade level prepared a manual for all teachers involved in that particular study. This manual consisted of a day-by-day plan of the specific classroom activities that would be carried out for the subject matter being taught. The exact manner in which the program was to be used was described and teacher materials for non-program instruction were specified. In this way, a degree of uniformity in the procedures being studied was accomplished.

In addition, a research coordinator checked with the teachers several times each week in the course of a particular study in order to insure that procedures were being carried out as had been planned.

(4) Student ability. Experience in the Baldwin-Whitehall School System has indicated that differences between classes in intelligence levels and previous subject-matter achievement influence achievement from programmed instruction. This is so despite the often quoted claim that with the individualization offered by programmed instruction the relationship between student attainment and measured intelligence will be near zero. There are a number of factors involved in assessing this statement which have been discussed elsewhere (Glaser, 1963). As a result it has been necessary in the separate studies reported to control the classes compared on the basis of average intelligence and achievement levels, and comparisons to assess the effects of the independent variables studied have required careful matching of class means in order to draw appropriate conclusions.

(5) Testing procedures and ceiling effects. Crucial to the assessment of experimental effects are the measures employed of the dependent variable, student achievement. In assessing the outcome of conventional and programmed instruction, various measures can be used, each of which has particular characteristics. Three main types of measures can be distinguished, namely, program tests, teacher-made tests, and nationally-standardized tests. Program tests are achievement tests which the program publisher considers an adequate sample of student performance to measure the objectives taught by the program. Teacher-made tests are developed in cooperation with the classroom teacher and consist of items representative of the expressed educational objectives of classroom instruction. Nationally-standardized tests are those commercially available

achievement tests used by schools to assess their instruction and compare themselves with national norms. All three of these types were employed in the various studies reported. Where the program test was not considered an adequate test of overall classroom objectives or of the program itself, it was supplemented by a teacher test or a nationally-standardized test. When a nationally-standardized test was used, agreement was obtained from the teacher and school administrators that this test was an adequate measure of their own course objectives.

If, in assessing experimental variations in the classroom, a definitive test is established to indicate mastery of the course objectives, then the objectives of instruction are to teach so that students attain such mastery. This means that in successful instruction many students will obtain perfect scores and the distribution of scores obtained for a class will be skewed with a ceiling imposed by perfect test performance. If two different instructional treatments are given to two different groups and both groups show many students with near perfect test scores, the problem is to distinguish which treatment represents the more effective instruction. Factors other than student achievement must be considered, such as time taken to attain mastery, etc. If achievement is the measure of concern, then the percentage of students obtaining a perfect score, the average level of mastery, or the gain in mastery from pre- to posttesting can be used. A question might always remain, however, with respect to how much more knowledge would have been exhibited by students if the test did not have a mastery ceiling. For example, if the objective of a course of instruction is to teach students addition and subtraction with single-digit numbers, a mastery test would measure just that skill, addition and subtraction with single-digit numbers; however, it is justifiable to ask to what

extent students can extrapolate and transfer their knowledge to two- and three-place numbers. The tests employed in the studies reported here are, for the most part, tests with mastery ceilings and were used to assess the attainment of specific mastery objectives. Sometimes tests of more general objectives were employed which did not display ceiling effects. There were usually nationally-standardized tests which are constructed so as to give a wide distribution of scores.

(6) Extrapolation of laboratory findings. As has been indicated, a number of the studies reported involved variables suggested by laboratory experiments. In general, the direct extrapolation of a laboratory variable to actual instructional practice in intact classes runs many risks. One is that in group experiments in the laboratory the differences between experimental and control groups are often obtained under stringently controlled laboratory conditions, and it can be expected that an effect of small magnitude under such control conditions will be attenuated in the conditions of the practical classroom. For the most fruitful interaction between the laboratory and instructional procedures in the classroom to take place, a research and development sequence is required which passes through fundamental laboratory research, through development, through design and proving and field tryout (Gilbert, 1962; Glaser, in press). Another aspect of extrapolation is that it is likely that variables found to be significant in group experiments have a higher probability of being attenuated in actual practice than have effects that have been replicated with individual subjects (Sidman, 1960). In keeping with the intact classroom approach, however, some of the experimental variations reported here represent extrapolations of variables that have been found in the laboratory to show significant effects between groups of subjects.

The Nature of the School System

The studies reported here have been carried out in the Baldwin-Whitehall Public Schools, situated in a suburban residential area contiguous to the City of Pittsburgh. The population of the area represents a cross-section of the metropolitan Pittsburgh area, ranging from skilled mill and industrial workers to executive and professional types. The school system consists of one high school, two junior high schools, and 12 elementary schools (kindergarten through sixth grade). The total student enrollment during the 1962-63 term was approximately 8000, with a classroom teaching staff of approximately 375.

First Grade

Introduction to Numbers. In the beginning weeks of first grade an exploratory study was conducted to determine how well very young students could work independently on programmed instructional materials. The program used taught the student to write and recognize the numbers from 1 to 10, to understand the concept of the number, and to recognize the differences between the numbers. The students worked for about 20 minutes each day. A two-part pretest measuring what the program was designed to teach was given prior to beginning the program. The first part required that the student respond orally to the printed numbers, that is, say them out loud, and copy the numbers; this was the easier section of the test. The second part, which was more difficult, required that the student write numbers as they were dictated and also count objects and write the number. The same examination was given as a posttest. Each student was permitted to progress through the program at his own pace, and Figure 1 shows the distribution of the number of days taken by 121 students to complete the program. The very wide distribution of completion times, from 8 to 25 days, indicates that the slowest student spent three times as much time

as the fastest on the program. Figure 2 shows a scatter diagram of time to complete the program and final score, and indicates that there was little correlation between these two factors for this program. Figure 3 shows the scores on both sections on the pre- and posttest. Pretest scores on Part 1, the more simple tasks, show that most students knew this part of the material before beginning the program. Fewer students knew the material covered by the second part of the pretest, and the posttest scores for this more difficult test indicate that they were successful in learning these tasks during the course of the program. The result of this exploratory study indicated several things: (1) the program was a reasonably efficient teaching device, (2) variation in rate of learning among students is extensive, and (3) systematic pretests show that many students know the subject being taught and some few students are not ready to learn it.

Following this exploratory study, the introduction of additional programmed materials was planned. The schedule of teacher instruction and programmed materials for the six classes involved in the remainder of the first-grade study is shown in Figure 4. The underlined portions indicate periods during which programs were in use.

Addition and Subtraction. Work with the addition and subtraction program was designed to observe procedures for coordination of teacher instruction and programmed self-study activity, and one of the purposes of the study was to compare three different program-teacher combinations. The program taught single-digit addition and subtraction facts, and the classes were scheduled for two 20-minute periods of arithmetic instruction per day, one period in the morning and one in the afternoon. In the first combination (Group T-P) students received initial instruction in a particular addition and subtraction

topic from the teacher during the first daily session, and practice and review of the same topic from the program during the second session. A second group (Group P-T) received initial instruction from the program, followed by practice and review under teacher direction in the second session before going on to new material. The third group (Group P-P) received only the program during both daily sessions for the first half of the experiment, working daily on the program unit until it was completed; following this the teachers reviewed all of the addition and subtraction facts during both daily sessions for the remainder of the experiment. For simplicity this third group was designated the prior-program group. During the course of the study, each teacher was given a manual outlining the procedures to be followed in order to insure that all three groups received approximately equal treatment during the teacher instruction sessions.

Intelligence data and pre- and posttest scores for all groups are presented in Table 1. Since the P-T Group was lower than the others in mean intelligence, and all three groups differed in pretest performance, statistical analyses of achievement were made for gain scores rather than posttest performance. At the end of the study the T-P and P-T Groups, who were initially higher on the arithmetic pretest, scored significantly higher on the posttest than the P-P Group, but the three groups were equal in the amount of learning gained during the course of the program. Pre- and posttest data for the two groups with daily teacher-program alternation are shown in Figures 5 and 6. Figure 5 indicates that a number of students in Group T-P demonstrated mastery or near mastery on the pretest and that on the posttest practically all students indicated mastery. Figure 6 indicates that few students in Group P-T were proficient in the subject matter on the pretest and that correspondingly

relatively fewer students attained mastery of the subject. Data for the group that completed the entire program and then received teacher instruction are shown in Figure 7. Pre- and posttest distributions show that students in these classes were successful in learning from the program. The interim chart shows arithmetic attainment on the same test given immediately following the program, before teacher instruction began. These interim scores point out that this program in itself did not teach as effectively as could be expected, and that while some students attained mastery with it, the role of the teacher insured that many more students achieved subject matter proficiency.

This study indicated that (1) some programs are not as efficient as should be expected, and (2) that different types of teacher-program combinations appeared to make little difference in student gain, since the gain scores were the same in the three groups. However, the effects that were being studied, that is, different program-teacher arrangements, may have been largely attenuated by the fact that the program used needed revision to become a more effective instructional tool. Since this was so, the program may not have articulated well with the teacher as an introductory or review device, and the teacher found it necessary, for some students, to compensate for program inefficiencies.

Time-Telling. The time-telling program used was a teacher-student programmed sequence in which the teacher worked with the students through the program. The program was an experimental one developed at the University; it had received a number of tryouts and revisions but was not fully perfected. It taught students to read a clock to the one-minute interval, and was constructed so that the teacher worked with the students as a group for approximately 20 minutes a day for 14 days, teaching intensively a topic which is usually taught only intermittently throughout the first and second grades.

During the course of the program students worked with booklets of flat, two-dimensional clocks, and were tested on these paper clocks and on real, Western Union-type wall clocks. The posttest results for a group of above average students (mean I.Q. 112.39) are shown in Figure 8. The solid lines show performance on the program clocks, and the dashed lines show performance on the real clocks. Seventy-seven percent of these first-grade children could respond correctly with only two errors to the paper clocks and 51% performed at this level with the real clocks. On a pretest prior to the program, none of the children got more than two items correct on the real-clocks test.

Several things were apparent from these data: (1) Such a program may be useful for teaching difficult special skills like time-telling, and for teaching on an intensive basis subjects that are usually taught more incidentally. A similar program might also be useful for such things as using a ruler. In these matters it is a moot point whether the concept should be taught before, after, or at the same time the skill is being taught. (2) The program was not as successful as could be expected, and revisions are necessary to increase its proficiency. One reason for not achieving greater success may have been that the program was paced, and a more individualized program with branching may be required. (3) Performance data showed that there was not sufficient transfer between program clocks and real clocks. At this age level learning can be very specific, and transfer must be recognized and built into the program.

Figure 9 shows the differences in responding to a real clock by a group of average students (mean I.Q. = 101.75, s = 10.34) and a group of above average students (mean I.Q. = 112.39. s = 14.51).[4] The mean score of the higher I.Q.

[4] As measured by the California Short-Form Test of Mental Maturity, Pre-Primary.

group was significantly higher ($t = 2.21$, df/55, $P < .025$) on the posttest, indicating that the program as built was sensitive to I.Q. differences.

Two weeks after completion of the program the time-telling tests were repeated as retention tests. Figure 10 shows posttest and retention data for the program clocks, and Figure 11 shows the same data for real clocks. In each case there was only a slight decrease in student attainment even though this subject was taught rather intensively over a short period of time.

With the time-telling program a further study was carried out on the effect of varying the distribution of work on the program. One group worked with it for one 20-minute period a day, another group for two 20-minute periods a day, and a third for one 20-minute period every other day. There was no difference in final performance between these groups. Also studied was the effect of prefamiliarization with clocks and time-telling games presented by the teacher before taking the program. Such prefamiliarization had no influence on learning from the program. However, additional classroom practice with the teacher after completion of the program was helpful in further improving learning.

Fourth Grade

At the fourth-grade level, programs in multiplication and division facts, introductory fractions, and spelling were used to assess (1) the effect of intelligence upon learning from programmed instruction, (2) the use of programming for review and acceleration of arithmetic learning, and (3) the effects of classroom surroundings upon learning from programmed materials.

Multiplication and Division. Programming has been recognized by educators and psychologists as one possible means for diminishing the dependency of learning upon intelligence as presently defined by standard tests. This is assumed for several reasons: (1) Since most intelligence tests have a speed

component and most group-paced learning situations maximize individual differences in speed, the relation between the common speededness components would contribute to a positive correlation between the two. In contrast, most programmed instruction is self-paced and would minimize the effect of learning rate and tend to lower the correlation of learning achievement score with a speeded intelligence test. (2) In a well-administered program the students have mastered the prerequisite behavior for taking the program. This should have the effect of reducing individual differences and lowering the correlation coefficient. (3) If a program is an effective instructional procedure so that more students achieve mastery than with other instructional procedures, the range of scores is reduced and consequently the size of the correlation coefficient is decreased. The first study in the fourth grade assessed the extent to which learning from a self-paced program was influenced by the intelligence of the learners.

The multiplication and division program used for this purpose taught the basic multiplication and division facts through operations with two-place numbers. Six classes were given the program over a period of six weeks. In a typical week the student worked at his own pace through the program during 45-minute work sessions on Monday, Tuesday, Thursday, and Friday. The programmed material was divided into ten sections, and as the student completed a section he was given a written test. If he failed to achieve a score of 70% on any section he was required to go through the program again and pass a retest at the 70% level. Wednesdays were set aside for teacher instruction periods, and at this time, following a prepared manual, the teacher presented review and practice materials relevant to the parts of the program which most students had completed.

As a result of the self-pacing procedure, some students finished the program before the end of the allotted six-week period, while others were unable to complete it within the designated time limit. Students finishing early were provided with enrichment materials to be used during the sessions when others were still working in the program. The slowest students in each class, for whom not enough class time was allotted for completion of the program, were required to take the program home several nights during the last two weeks and work on it there as well as in the classroom program sessions.

Figure 12 shows pretests and posttests on 25 multiplication and division items from the arithmetic subtest of the Stanford Achievement Battery plus 15 additional but similar items. Pretest scores reveal that some students already knew much of the program while others may not have mastered the prerequisite behavior necessary to profit greatly from the program. However, the posttest data indicates that this was a fairly effective program.

For the 173 students for whom complete data were obtained, the correlation between intelligence as measured by the Otis Quick-Scoring Mental Ability Test and the previously mentioned posttest was .19. On a second posttest consisting of items taken from the program, the correlation between I.Q. score and achievement was .20. Taken alone these correlation coefficients show that intelligence as measured accounted for very little of the achievement following the program. However, it is known that the restriction of range which comes about when a test has a ceiling lowers the size of the correlation coefficient. This was so in the present case, since many of the students achieved perfect scores on the test. Since the size of the correlations may have been affected by the test ceiling restrictions, a second type of analysis was performed. Above average and average I.Q. groups with similar pretest performances were selected from the original group, and their posttest achievement levels were compared.

The above average group (N = 28) had Otis I.Q. scores between 120 and 140. The average group (N = 29) had I.Q. scores between 90 and 110. Both groups had multiplication and division pretest scores of below 20, which was less than 50% of the 40-item test. Comparison of the differences in posttest mean scores and in mean gains showed no differences between the two groups. The difference in intelligence between the groups apparently had little effect upon final achievement or the amount of learning which took place with this program as measured by the final test.

Fractions. After completing the multiplication and division program, the six classes were divided into three groups of two classes each. Two of the three groups were presented the arithmetic curriculum that was currently being followed by the school system, using regular classroom instruction for the remainder of the year. The teachers followed lesson plans outlined in a manual especially constructed for this study. The only difference in treatment between these two groups was that one of them permitted low-achieving students to use the program as a review tool one month following the program. The third group was given programmed instruction in fractions soon after completion of the multiplication and division program. This constituted an acceleration of the arithmetic curriculum, since instruction in fractions is normally not a part of the fourth-grade subject matter in the school participating in the study. The three groups, the non-review group (Group NR), the group using the program as a review tool for the low-achieving students (Group R), and the accelerated fractions group (Group F), were matched on mean I.Q. and multiplication and division achievement. Figure 13 shows the fractions pretest and posttest distributions for Group F. Figure 14 shows the posttest score distrubutions on the fractions test for Group F as compared with the

combined data for the two groups that followed the regular curriculum. Figures 13 and 14 would indicate that the fractions program was an effective teaching device.

At the end of the year, all three groups were given two tests in multiplication and division, Test MD-A which was essentially the arithmetic section of the Stanford Achievement Battery plus some additional items, and Test MD-B which was more oriented toward the program. The mean scores on these tests are shown in Table 2. Both the means and standard deviations presented for the non-review, review, and fractions group indicate no difference between the three groups. It can be concluded, then, that in this case the review provided no additional learning, and that the fractions group, which spent less time on multiplication and division, reached an achievement level equal to that of the other two groups. On the fractions test itself the fractions group did much better than the other groups, as could be expected. In general, it seems that the eight weeks of arithmetic periods spent by the fractions group in learning additional advanced material, necessarily taking away learning time from the usual fourth-grade arithmetic activities in which the other groups were engaged, did not detract from learning or retention of the usual fourth-grade arithmetic topics. This strongly suggests that extension of the curriculum with programmed material produced a significant amount of additional arithmetic learning without being detrimental to the learning of material in the standard curriculum.

Spelling. Also at the fourth-grade level, a spelling program was used to evaluate the effect that variations in classroom environments have upon learning. It is an increasingly common practice in present-day school systems to teach various subjects in classrooms specifically designed for instruction in certain subject matter areas, e.g., science rooms, language rooms, etc.

A major reason for this is that the special equipment needed may be too bulky or expensive for the regular classroom, but it is also possible that the uniqueness of the surroundings in itself may facilitate the learning which takes place in such a room. This assumption is supported to some extent by laboratory investigations of human learning. The theoretical explanation involved is the concept of interference which postulates that if, in the course of learning, the incidental stimuli present are already associated with responses other than the ones being learned, these older associations tend to interfere with the new ones being made. This associative interference hypothesis suggests that when several different subject matters are being taught in the same classroom, the responses appropriate to one subject matter may be associated with classroom stimuli in a way which interferes with learning responses to other subject matters. Therefore, the unique surrounding stimuli present in a classroom used for only one subject matter may diminish interference effects upon learning. Also, once a response has become associated with these incidental stimuli, the presence of such stimuli may facilitate retention of material learned if a test is given in their presence, since they would tend to evoke the appropriate associative responses rather than interfering responses.

Although this hypothetical explanation of possible effects of incidental stimuli upon learning is quite crude and general from a rigorous point of view, it does present a possible extrapolation of existing theoretical descriptions of the learning process to educational practice and suggests an exploratory experiment. Consequently, a study was performed with fourth-grade programmed spelling to explore the hypothesis that incidental classroom stimuli will facilitate or inhibit learning. It was predicted that a group receiving both spelling instruction and spelling tests in a special room in which no other subject matter

was taught would demonstrate more learning than a group given the same instruction and testing in a room used by that group for learning other subject matters as well. A third group, receiving spelling in a special room but all spelling tests in the usual classroom, was used to determine the effect of surrounding stimuli upon test performance alone. It was predicted that test performance of the latter group would be lower than that of the group which received all tests in the special room.

Six classes in the fourth grade, divided into the three groups described above, received programmed instruction in spelling, using a programmed text which taught 354 new words. All groups, plus a control group receiving traditional instruction, were equivalent in terms of mean I.Q., mean pre-instructional spelling achievement level (measured by the spelling subtest of the Stanford Achievement Battery), and mean level of general academic achievement (measured by the Stanford battery median score). All experimental groups worked on assigned frames during scheduled 20-minute class periods on Monday, Tuesday, and Wednesday of each week. Faster students who finished in less than the three 20-minute periods allotted were given individual spelling enrichment tasks by the teacher. Students who could not finish in the allotted time were given extra time to insure that they would be able to participate in the teacher-directed enrichment periods. Statistical analysis of scores of the program groups on end-of-year tests showed that all three groups were equivalent on the program tests and the spelling subtest of the Iowa Test of Basic Skills, indicating that in this study different environmental stimuli had no effect upon learning.

Figure 15 shows the pretest and posttest distributions for all program groups combined on the spelling section of the Iowa Test of Basic Skills and on a test of the words in the program. Pretest scores indicate that a number of the children knew what was to be taught before beginning the program. However, this figure also indicates that the program was quite effective. Figure 16 shows a comparison of the program groups and the controls on the more general Iowa test, and Figure 17 shows the comparison of the two groups on the test constructed specifically to measure learning of the program words. These figures indicate that on the nationally-normed, more general test there was little difference between the program group and the non-program group, but on the specific program test the program group was more successful. The time allotted to all spelling instruction for the two groups was equal. In the light of this it appears that the program group was able to learn more spelling by being required to learn more in the same period of time, and this acceleration did not detract from their learning to spell words usually taught at this level.

Seventh Grade General Science. In the seventh grade, a general science program was used to determine the extent to which enrichment activities following programmed instruction facilitated student achievement. The enrichment activities consisted of filmstrips and movies, laboratory experiments and demonstrations, group discussion, assignments in source materials, and individual and group projects. Two groups, equivalent in I.Q. as measured by the Otis Quick-Scoring Mental Abilities Test, science achievement as measured by the Science subtest of the Stanford Achievement Battery, and general achievement as measured by the battery median score of the Stanford Achievement Battery, were given paced programmed instruction through six units of a general science program. There were 63 students in each group. Each group spent the same

amount of time on each science topic. A long-enrichment group (LE), however, was required to complete program units in a shorter period of time, allowing more time for enrichment activities. A second group (SE), the short enrichment group, spent more time on the program and less in enrichment. Over the 66 school days during which the study took place, Group SE spent 19 days in enrichment activities. Group LE spent 30 days, or 11 more class periods in enrichment activities than Group SE, and consequently spent 11 fewer days working on the program than Group SE. At the end of the semester measures of science achievement were administered to both groups. Figure 18 shows the pre- and posttest scores of both groups on the program tests. Critical ratio tests showed that the group means did not differ significantly on either the pretest ($CR = .71$, $P > .05$) or the posttest ($CR = .80$, $P > .05$), indicating that the treatment variations in amount of enrichment provided during the study had no effect upon an overall measure of the amount of programmed material retained at the semester's end.

A further analysis was performed using the scores of Groups SE and LE on the six program unit tests that were administered during the course of the program immediately following the end of the enrichment period for each topical unit. Significantly higher mean scores were obtained for Group LE on three of the unit tests. The longer enrichment period apparently had a facilitating effect on achievement immediately following learning, although over a longer period of time the effects of the two kinds of enrichment were not evident.

A second study with seventh grade general science involved the evaluation of the effect of prefamiliarization procedures upon achievement following the use of programmed materials. It is assumed by many educators and psychologists that a period of familiarization with materials that are to be learned

facilitates subsequent learning. This is evidenced by the fact that most formal study plans contain references to an overview or motivation period in which the teacher initiates a new topic by familiarizing the student with the materials to be presented. Evidence from human learning research which indicates the importance of response learning prior to association learning, lends some support to this educational practice. This study compared a group receiving response familiarization (RF)(N = 47) with one that did not (N = 45). In the RF group, teachers presented an overview of the subject matter and required that the students take a spelling test, on the new terms they were to learn, prior to working with the program. Both groups devoted the same number of days to program work and to enrichment. However, part of the enrichment time for Group RF was used for overview and response prefamiliarization. Group means indicated that the RF group scored significantly higher on program and standardized posttests. Since Group RF also scored significantly higher on pretests of science achievement, an analysis of covariance was performed to adjust for this difference. This procedure showed no reliable differences between the two programs.

Ninth Grade

Algebra. At the ninth grade level a program in beginning algebra was used to study the effect of program use on above average and average I.Q. groups. Two groups of students with different mean I.Q.'s were given the algebra program (mean Otis I.Q.'s were 118.75 and 110.37). These two groups were compared on pre- and posttests with two matched groups receiving traditional instruction in algebra (mean Otis were 117.58 and 111.83). The programs were administered in the same manner to both experimental classes. Students proceeded at their own rate through each of the 30 units in the program, and as a student completed

a unit he was given the appropriate unit test. If a student failed to obtain a unit test score of 70 to 80% or higher, he was required to work through that unit again and take the test again. During one period of each school week every student attended a small-group review and discussion session conducted with students who were at approximately the same place in the program.

Final course achievement as measured by the Cooperative Mathematics Test, Algebra I, Form A (Educational Testing Service, 1962) is shown in Table 3. Statistical analysis of these data show that Group A1 scored significantly lower than the other three groups. There was no significant difference between the scores of Program Group 2 and Control 2 nor between Groups Control 1 and Control 2. On the basis of these data alone, it is possible to conclude that the program was less effective for the average student than it was for the brighter student. However, when the mean number of units completed for the two groups is compared, Program Group 2, the above-average group, completed a significantly greater number of program units. The mean number of units completed by Program Group 2 was 20.7 ($s = 5.1$) and by Program Group 1 was 18.4 ($s = 4.8$). In light of this, the conclusion that the program was less effective for the average group is explained on the basis of the ability of the above-average group to go through the program at a faster rate and cover more material in an equal amount of time. In a separate study of a group of 131 students, a correlation coefficient of .68 was obtained between the number of program units completed and the Cooperative Algebra Test score.

Implications

The broad conclusions that can be gathered are the following:

1. There is extensive variation in rate of learning among students when they are given the opportunity to proceed at their own rates with programmed learning materials.

2. Pretest scores show that many of the students know the subject being taught and some few students are not ready to learn it.

3. Different types of teacher-program combinations in several grades made little difference in student achievement.

4. Young children can be taught a subject intensively with little loss in retention (at least over the short time measured in this study).

5. The extent of the correlation between general intelligence and achievement as a result of programmed instruction depends upon the particular program involved. In general, intelligence appears to be related to the pace with which the student goes through a program.

6. Extension of the curriculum with programmed materials, necessarily taking away from time spent in conventional grade-level instruction, produced additional learning without being detrimental to the learning of materials usually taught at that grade level. In general, students required to learn more did learn more.

Most impressive in these studies was the wide variation in student rate of learning and the wide variation in student achievement prior to instruction. As a result, attention in the 1963-64 academic year at the Baldwin-Whitehall Schools is being focused on the individualization of instruction. Intensive pretests are administered prior to the beginning of instruction so as to determine the level of each child. In the course of instruction teachers are provided with enough clerical assistance and materials to immediately evaluate their students and keep them advancing. When programmed materials and other materials are used, students take a pretest on each teaching unit, and, if they reach a specified criterion score, can skip that unit and go on to the next. In this way repetition of already learned material is eliminated and students

can advance more rapidly to new learning. If a student does not have the knowledge necessary to begin a program, he is tutored or given special assignments to bring him up to the level necessary to enter the program. Using this procedure, many fourth-grade students have skipped through much of the spelling and multiplication-division programs, completed the fractions program (usually considered to be fifth-grade work), and will have completed decimal fundamentals by the end of fourth grade. Third-grade students are being given the spelling program previously given in the fourth grade, and although they progress through it more slowly, they are successful. It is also possible that some second-grade students will be ready for this program during the year. The introduction to numbers program is slated for use during the second semester of kindergarten rather than first grade this year. By the end of the year it is expected that in some classes students will be ready for work that is several grades ahead of their conventional placement and other students will have just completed, or nor quite completed, the work required at their normal grade level. A situation like this brings us a small step closer to our aspirations for individualizing the educational process.

Table 1

Means, Standard Deviations, and Mean Percents for Intelligence Data, Pre- and Posttest Scores of the Three Groups in Addition and Subtraction

				Total Test (90 Items)			Addition (45 Items)			Subtraction (45 Items)		
Group	N		I.Q.	Pre	Interim*	Post	Pre	Interim*	Post	Pre	Interim*	Post
T-P	44	$\overline{X}$	112.21	42.68		87.25	27.16		44.23	15.52		43.00
		s	14.33	26.06		11.82	13.33		.77	14.50		2.15
		%**	---	47		97	60		90	34		96
P-T	34	$\overline{X}$	103.59	21.56		62.65	15.65		33.79	5.79		28.85
		s	11.53	20.00		26.85	13.43		12.67	9.20		14.89
		%**	---	24		70	35		75	13		64
P-P	43	$\overline{X}$	112.37	32.33	54.40	78.35	23.26	33.53	41.14	10.93	20.86	37.21
		s	14.58	25.36	27.39	14.36	15.05	13.30	5.85	17.04	16.52	9.33
		%**	---	36	60	87	52	75	91	46	46	83

* Administered to P-P Group only.

** Percent Correct.

Table 2

End-of-Year Test Data for all Fourth Grade Program Groups

Group	N	MD-A 40 Items		MD-B 40 Items	
		$\bar{X}$	s	$\bar{X}$	s
NR	56	30.0	4.0	33.4	4.6
R	65	31.5	5.1	33.9	5.3
F	52	32.4	4.8	33.9	4.5

Table 3

Means and SD's of Program and Control Groups on Otis I.Q. test, and Cooperative Mathematics Test, Algebra I, Form A.

		I.Q.		Coop. Algebra Test 40 Items	
Group	N	$\bar{X}$	s	$\bar{X}$	s
Program 1	67	110.37	6.61	18.71	4.11
Control 1	60	111.83	10.47	22.57	6.48
Program 2	64	118.75	6.38	21.72	4.95
Control 2	26	117.58	5.84	22.04	6.47

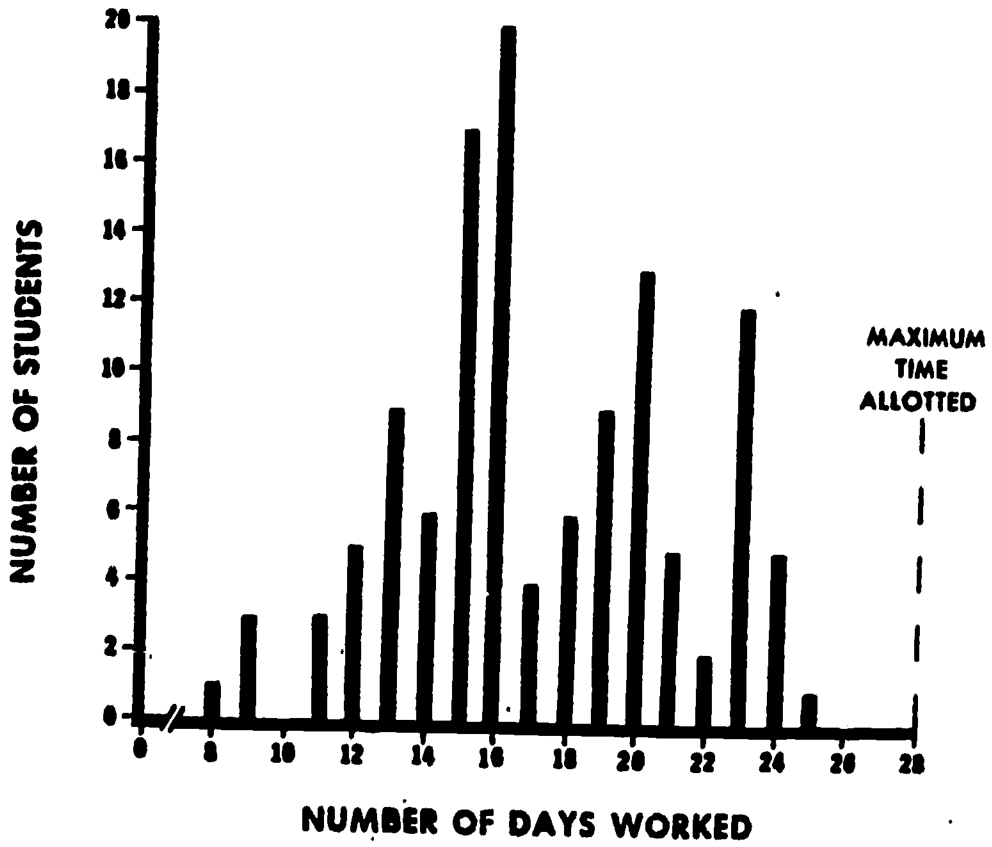

Figure 1. Number of days to complete introduction to numbers program (N = 121).

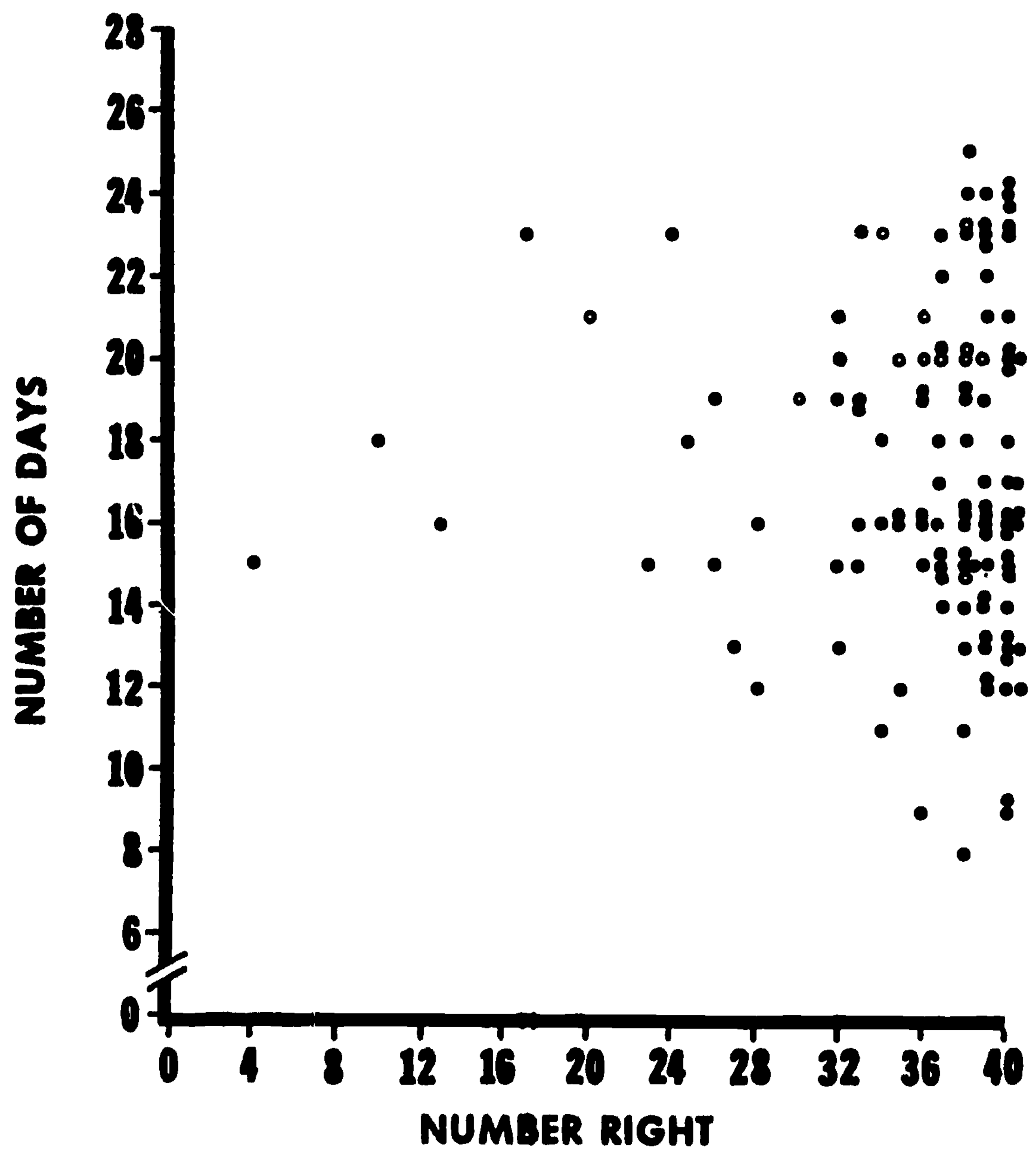

Figure 2. Scatter plot of final test scores and number of days to complete Introduction to Numbers program.

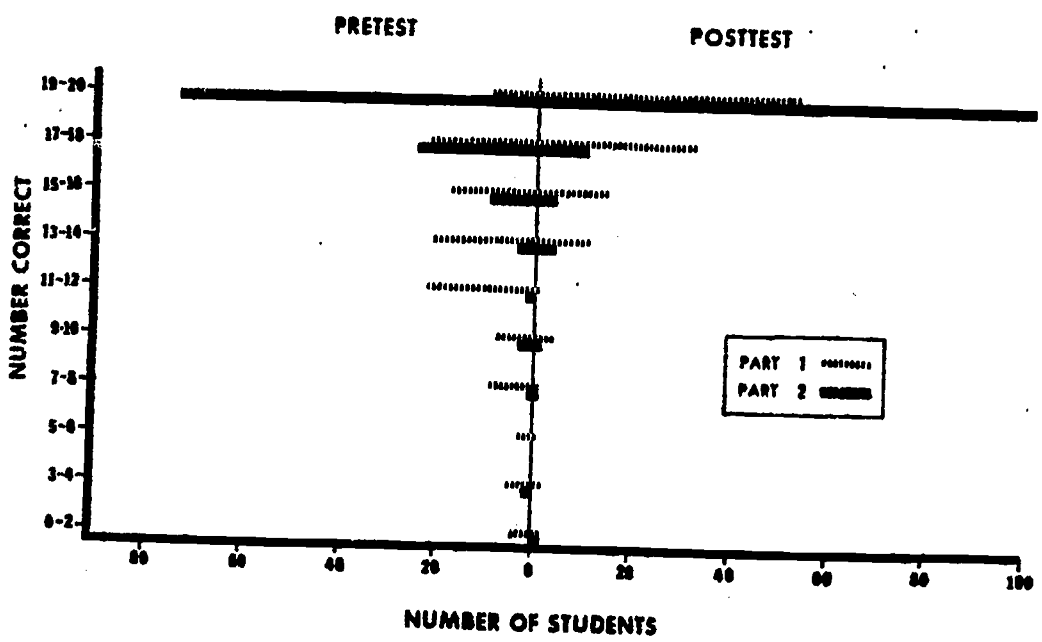

Figure 3. Pre- and posttest scores on introduction to numbers program.

Sept. 10 - Sept. 14	Teacher instruction: number readiness
Sept. 15 - Sept. 21	Practice with teaching machines
Sept. 24 - Oct. 19	Introduction to Numbers program
Oct. 22 - Nov. 23	Teacher instruction: number review, measurement
Nov. 26 - Dec. 8	Teacher instruction: addition and subtraction readiness
Dec. 10 - March 15	Addition and subtraction program and teacher instruction
March 18 - April 18	Teacher instruction: review of addition and subtraction, money, measurement
April 21 - April 29	Teacher instruction: counting by 5's and/or prefamiliarization for time-telling program
May 1 - May 20	Time telling: a teacher-student programmed sequence
May 21 - June 3	Teacher instruction: arithmetic review or time telling post-learning practice

Figure 4. Schedule for first-grade arithmetic.

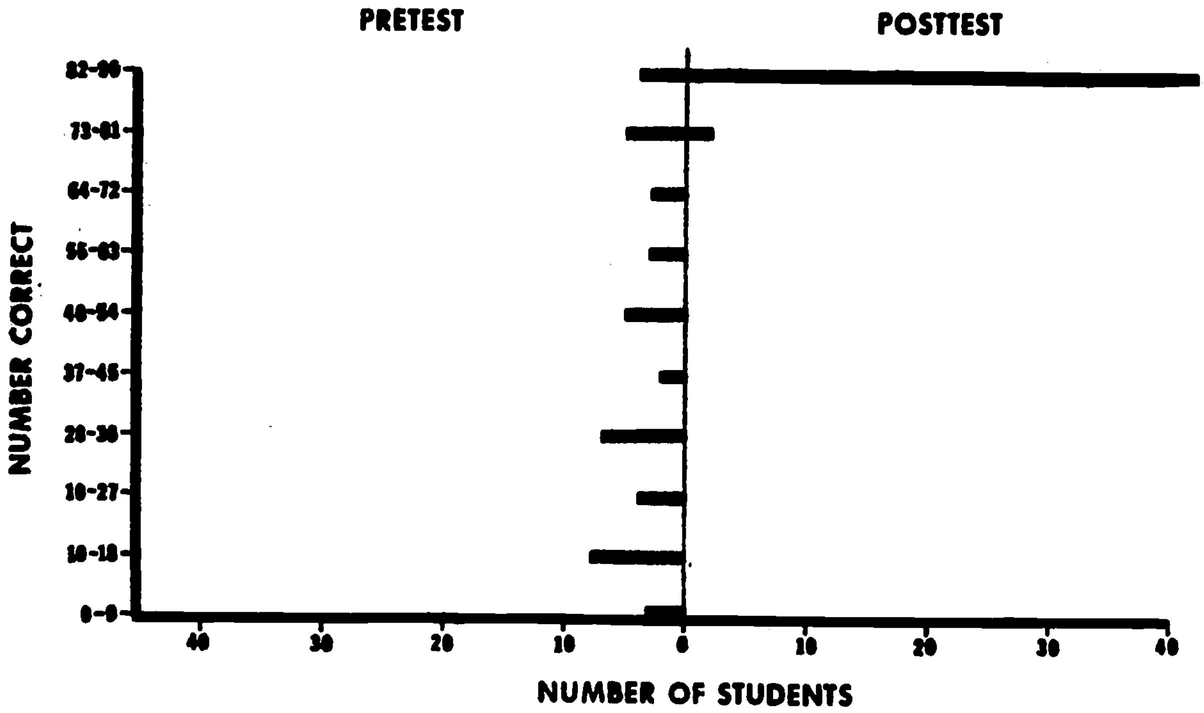

Figure 5. Pre- and posttest distributions for addition and subtraction, Group T-P (N = 44).

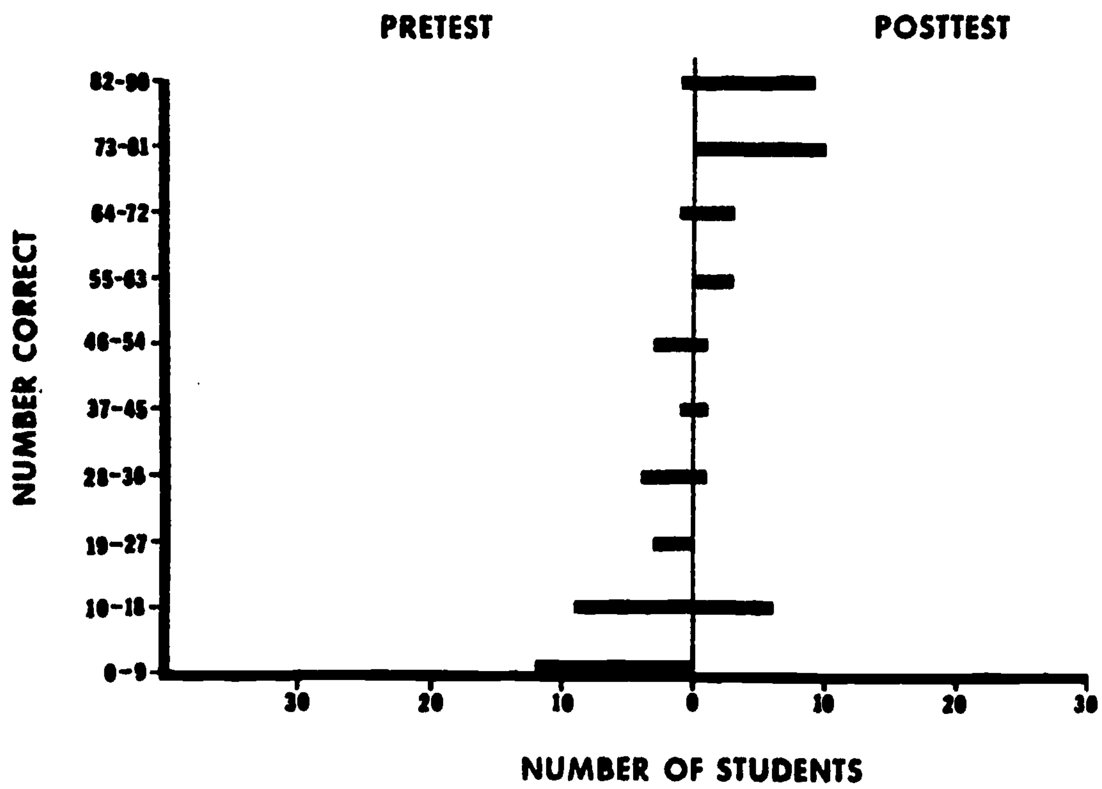

Figure 6. Pre- and posttest distributions for addition and subtraction, Group P-T (N = 34).

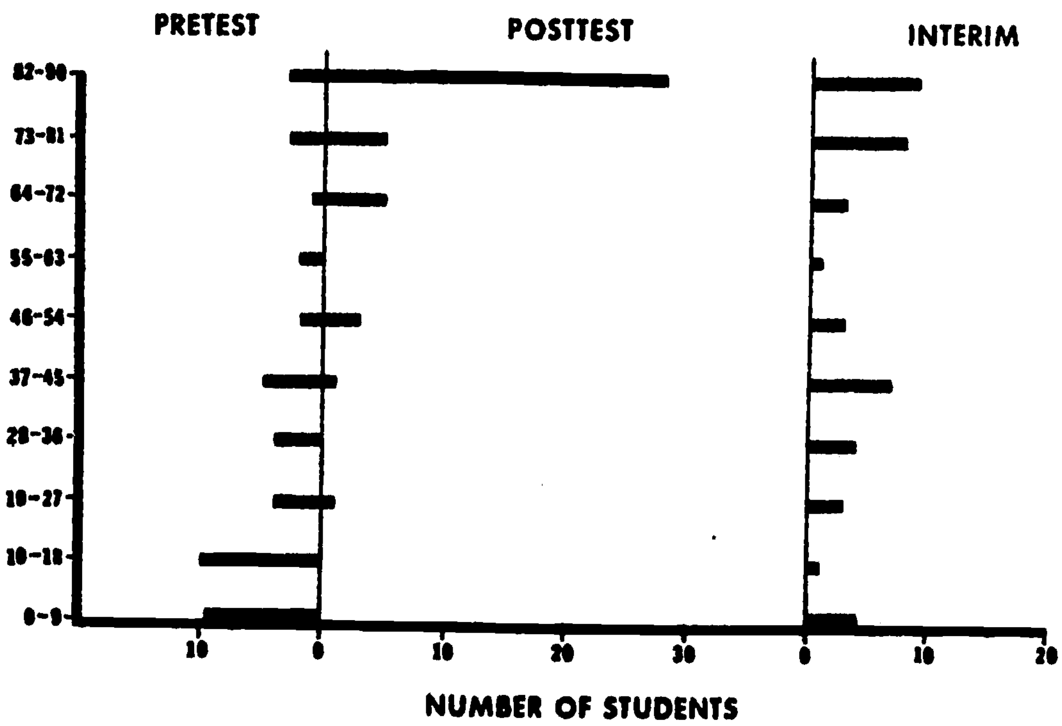

Figure 7. Pre-, post-, and interim test distributions for addition and subtraction, Group P-P (N = 43).

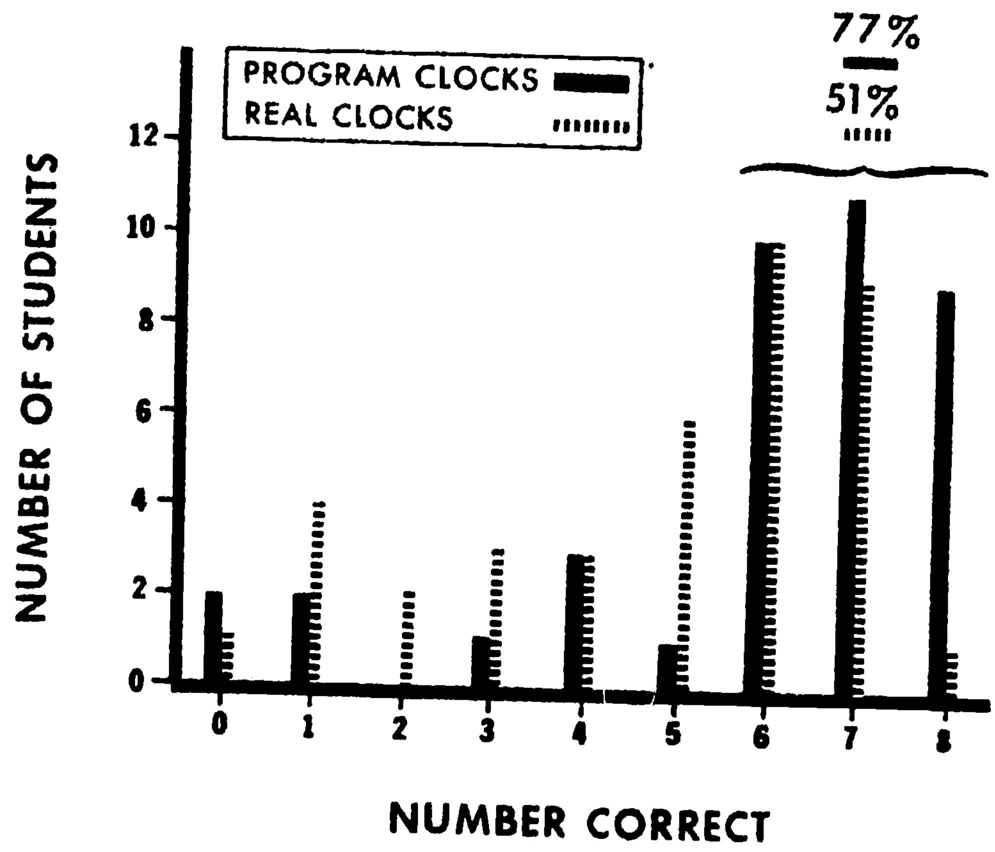

Figure 8. Posttest scores on program and real clocks (N = 39).

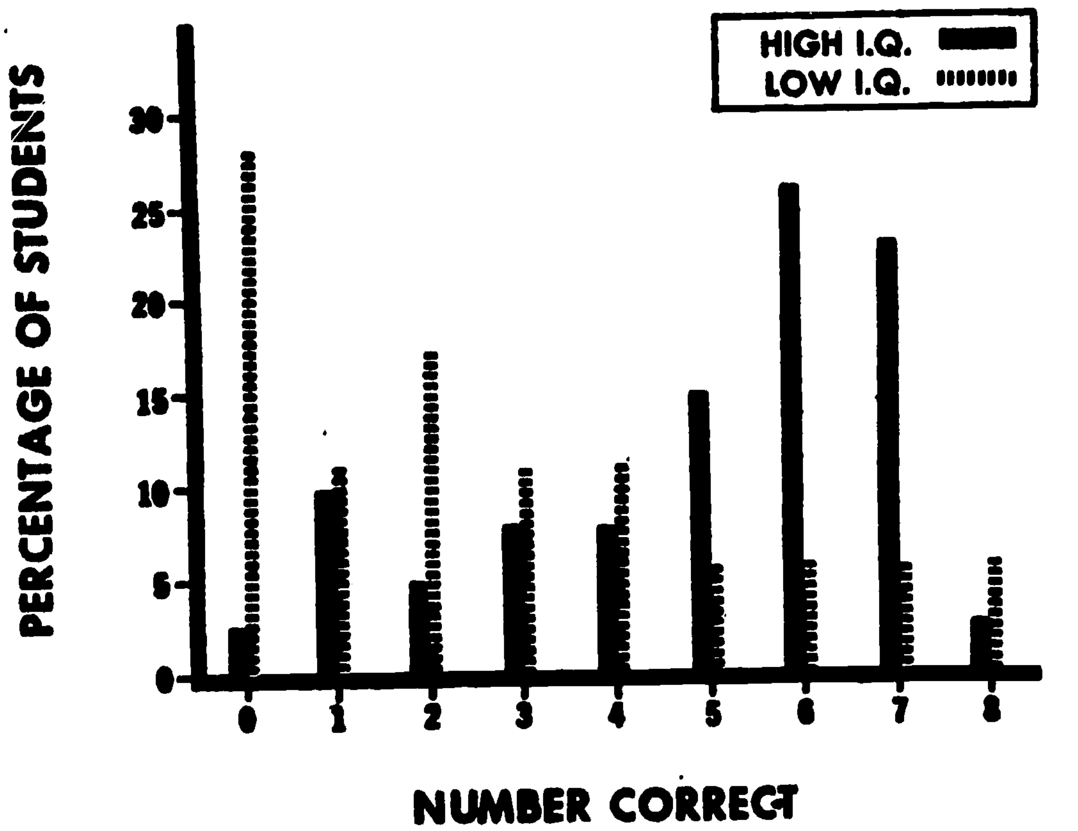

Figure 9. Posttest scores on real clock test by above average (Mean I.Q. = 112.39) and average (Mean I.Q. = 101.75) students.

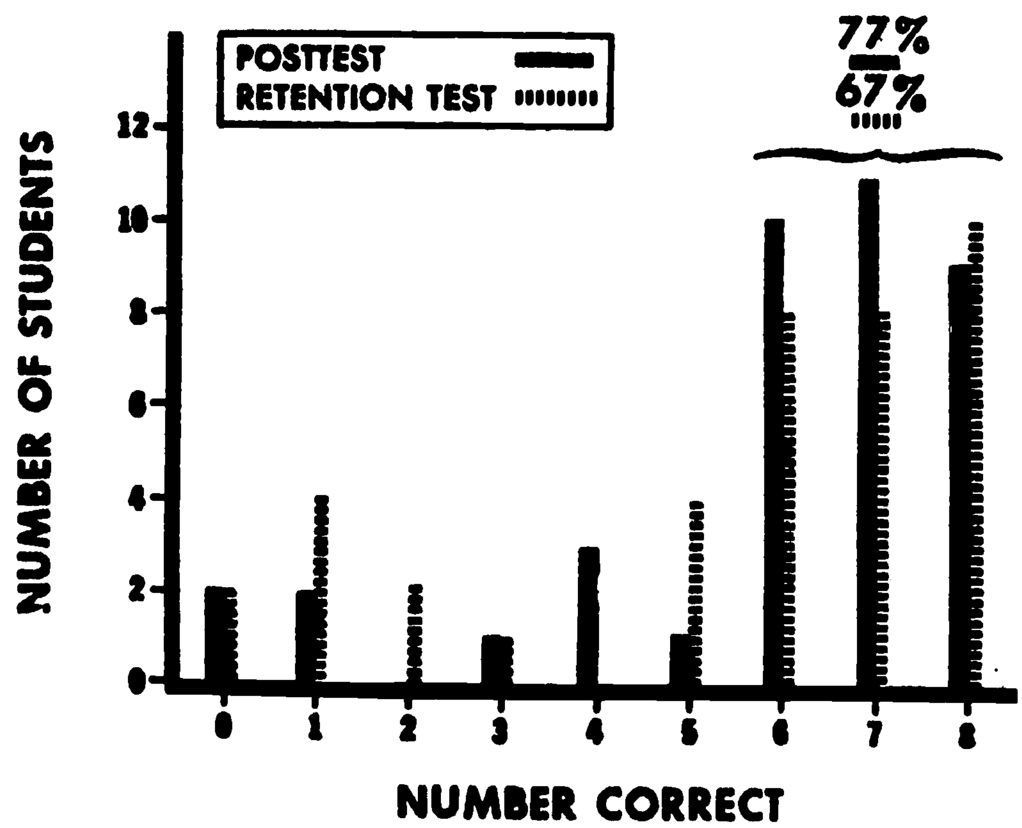

Figure 10. Posttest and retention test data for program clocks test (N = 39).

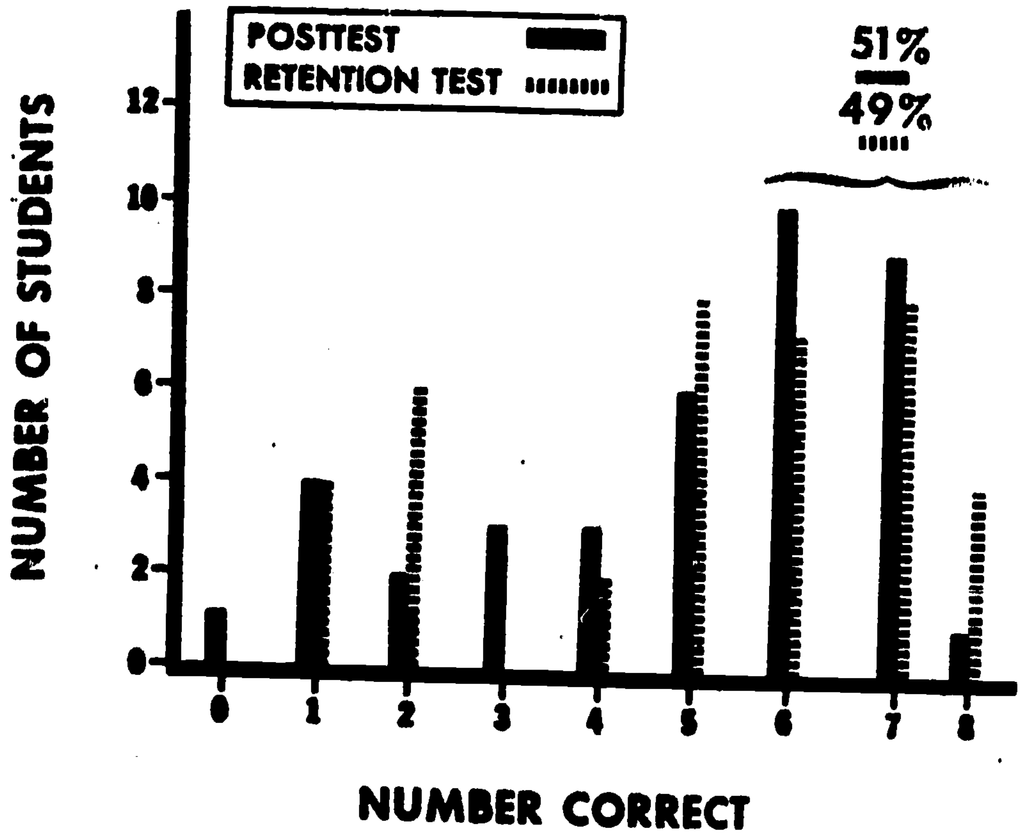

Figure 11. Posttest and retention test data for real clocks test (N = 39).

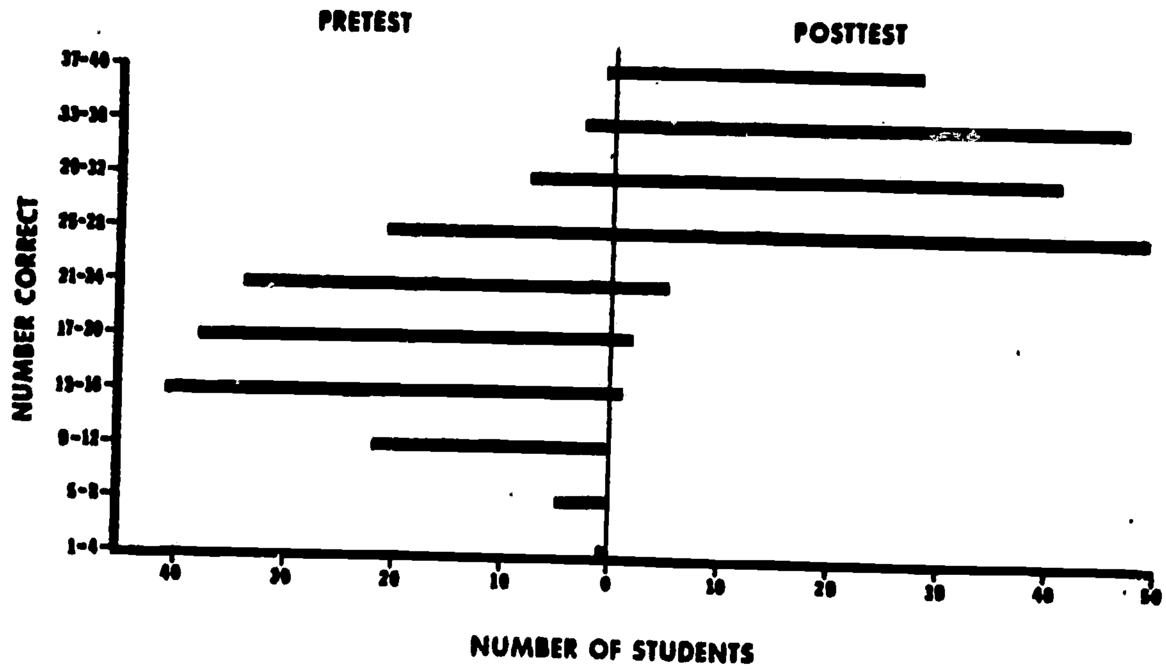

Figure 12. Pretest and posttest scores on multiplication and division (MD-A) test (N = 173).

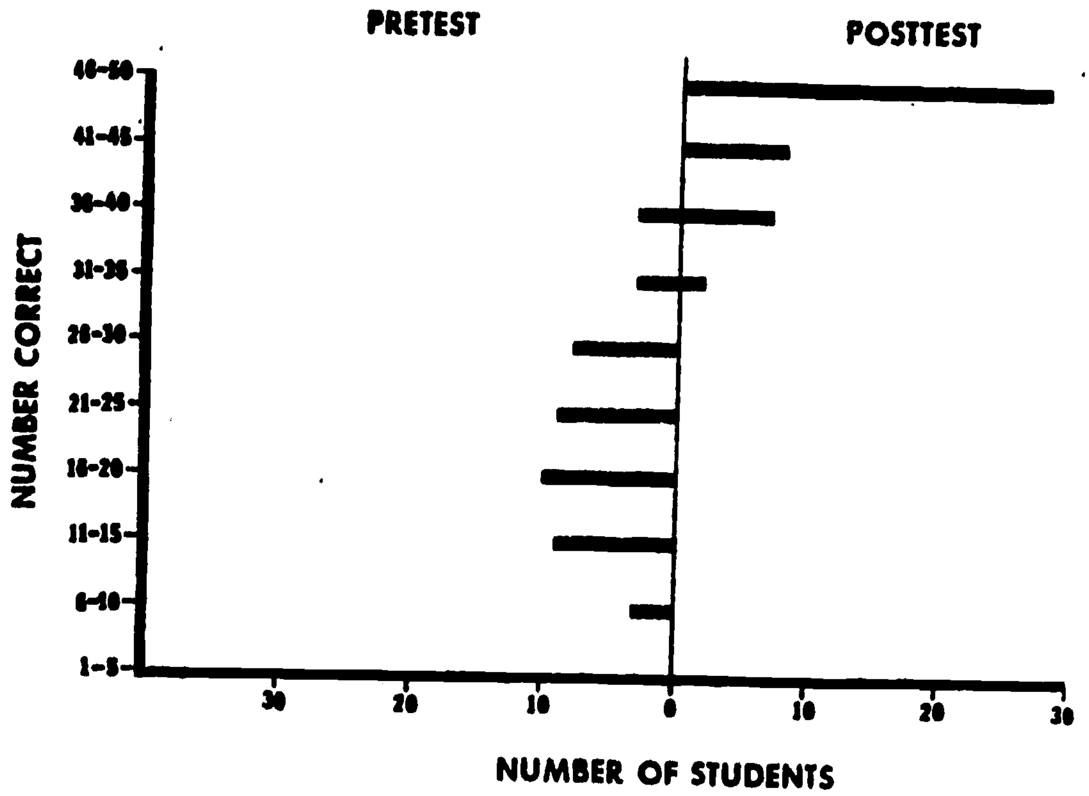

Figure 13. Pretest and posttest scores on fractions test (N = 45).

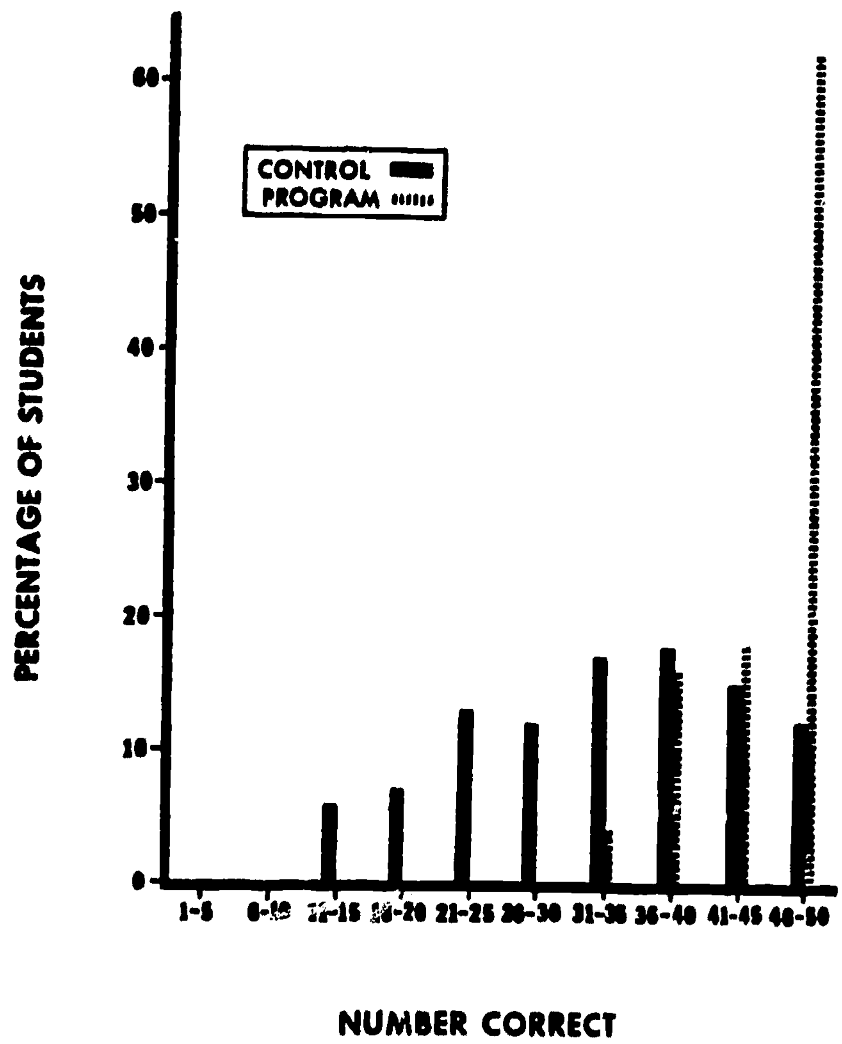

Figure 14. Posttest scores of program (P) and control (R and NR) groups on fractions test.

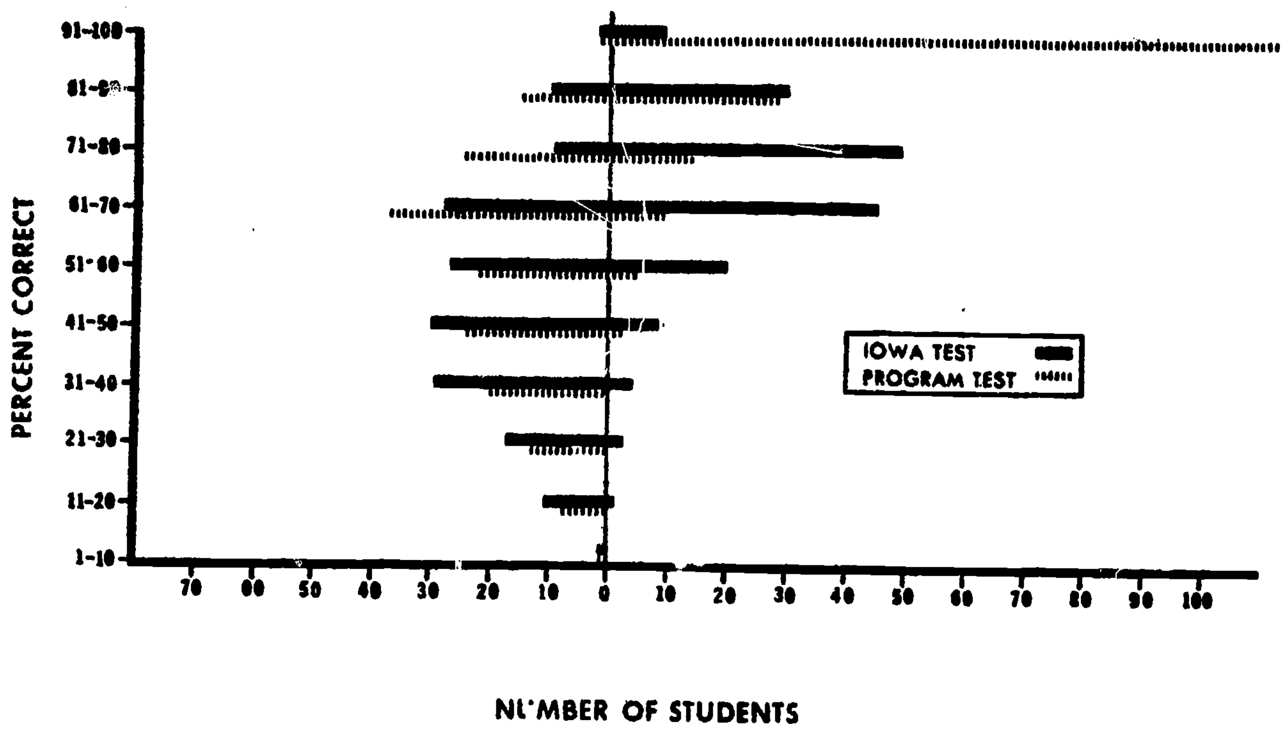

Figure 15. Pretest and posttest scores for all program groups combined on program test and Iowa Spelling Test (N = 169).

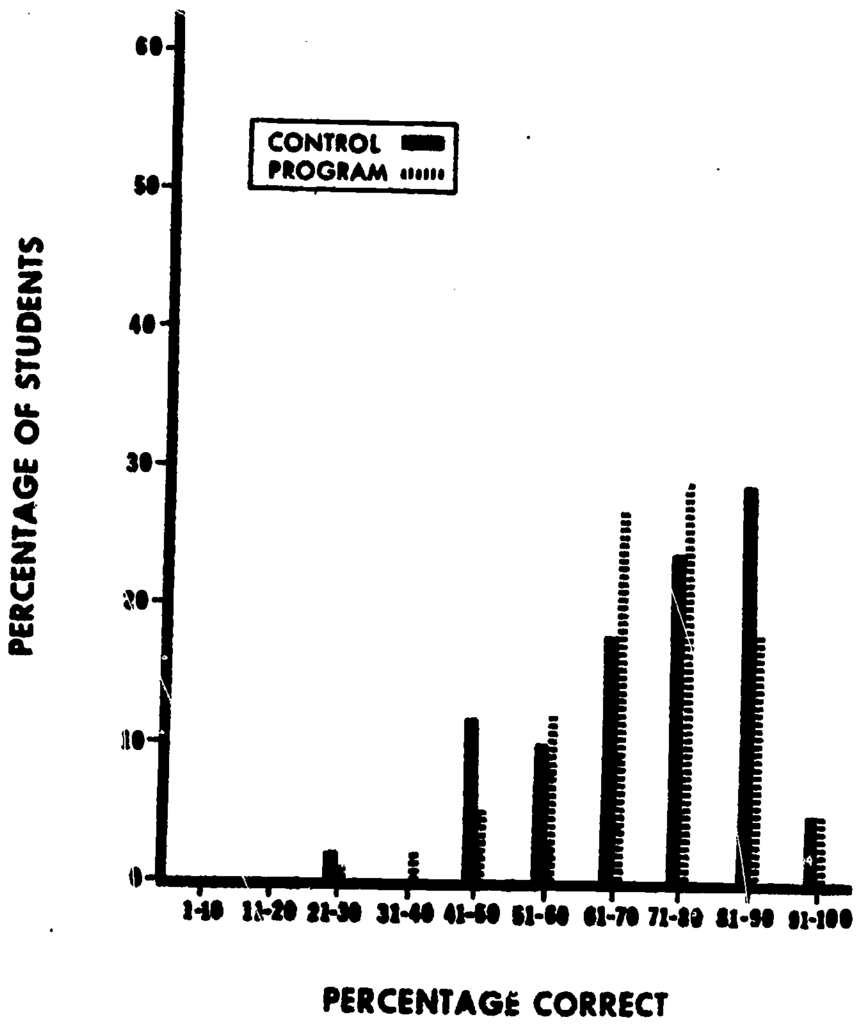

Figure 16. Posttest scores of program and control groups on Iowa Spelling Test.

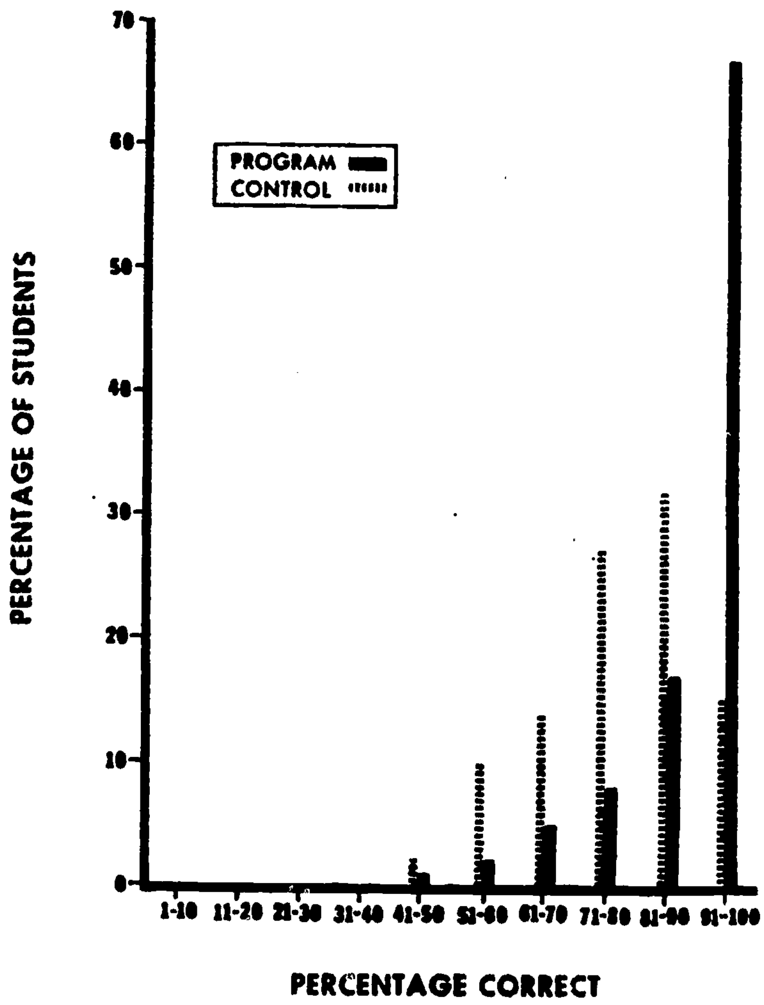

Figure 17. Posttest scores of program and control groups on program spelling test.

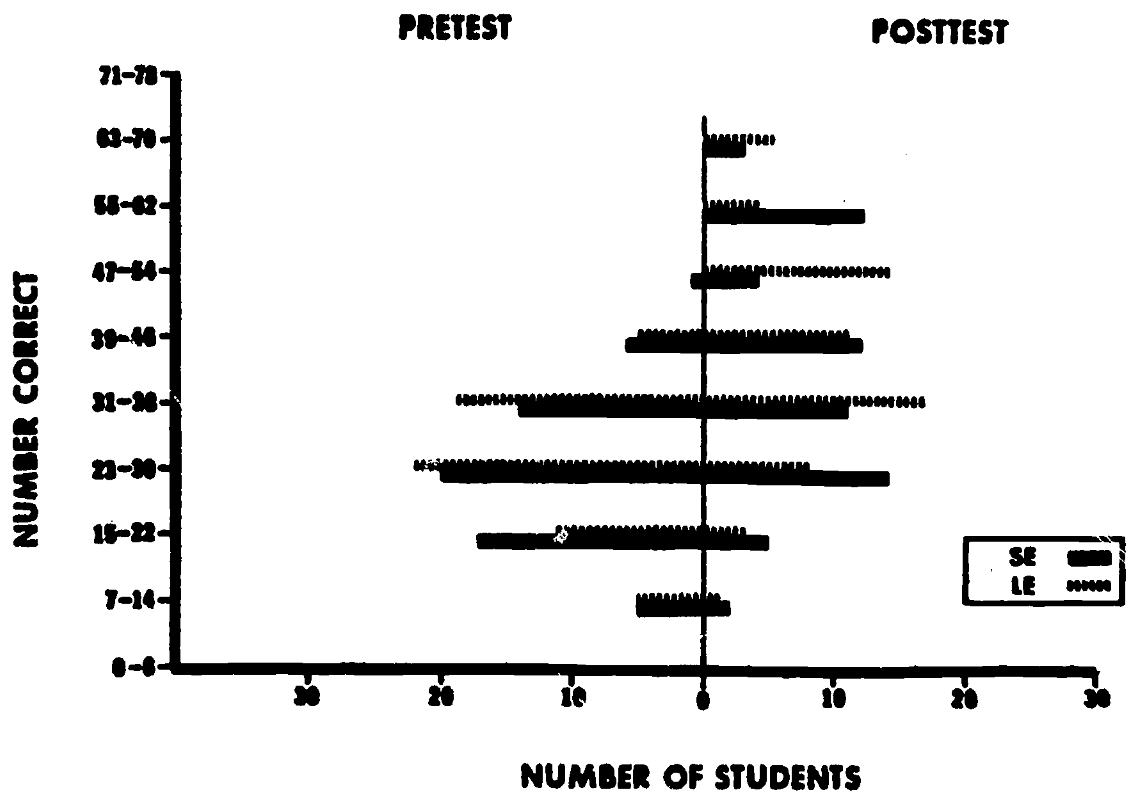

Figure 18. Pre- and posttest scores of Groups SE (N = 63) and and LE (N = 63) on general science program test.